Predicting Adolescent Response to School-Based Mindfulness:

A Machine Learning Analysis in the MYRIAD trial

eMethods. Participants, Procedure, School Characteristics, Measures, Analytic Approach

eResults. Supplemental Results

eTable 1. Predictor Definitions

eTable 2. Baseline Characteristics of Students by PAI Group

eFigure 1. Trends in Depression Scores in Every School in the MYRIAD trial

eFigure 2. QINI Curves

Methods

Participants and Procedure

We performed secondary analyses using data from the MYRIAD trial, a cluster-randomized controlled trial (RCT) with a two-arm parallel group superiority design (ISRCTN86619085; registered 03/06/2016). The two arms included the SBMT program which teaches mindfulness skills via psychoeducation, class discussion, and practices. The TAU control condition was included to compare the SBMT intervention to standard social-emotional learning (SEL) (e.g., covering topics such as relationships, sex education, and physical and mental health education) provided in secondary schools.[1] The trial commenced in 2016 and included 8,376 adolescents aged 11 to 13 at baseline, drawn from 84 schools across the UK (See **Table 1** for school and student characteristics). Both schools and students were broadly representative of UK secondary schools and their populations. Although 85 schools were initially recruited and randomized, one school withdrew after randomization but prior to intervention, following baseline assessment. Recruitment occurred in two waves during the 2016/2017 and 2017/2018 academic years. The process involved obtaining consent from schools, offering parents the option to withdraw their children, and securing assent from participating students. Eligibility criteria for participating schools included: (1) mainstream secondary status within the UK educational system (public or private schools), (2) employment of a permanent headteacher, (3) had not been deemed "inadequate" in their most recent official inspection to minimize implementation risks,[2,3] and (4) existing infrastructure and protocols for SEL curriculum delivery. Detailed information on the study design and methodology, as well as descriptions of the SBMT and TAU interventions are available in the published protocol[1] and its subsequent

update.[4] We analyzed data collected at baseline and post-intervention to capture intervention effects. See prior MYRIAD publications[1,3,5] for additional trial details.

SEL Across Study Schools

In the parent MYRIAD trial, "teaching-as-usual" (TAU) reflected the social-emotional provision already in place within schools, which varied across the UK. In England, this was most often delivered through Personal, Social, Health and Economic Education (PSHE), a subject with non-standardized implementation. To capture this variability, at baseline we collected structured information on each school's provision from a staff member responsible for the curriculum or school leadership. Provision was then rated against a set of 16 quality indicators developed through expert consultation and literature review, yielding an overall score (0–16) with higher scores reflection higher quality SEL provision. These assessments provided a standardized way of comparing schools within the UK context. In our dataset, there was no significant difference between UK countries in these SEL scores, $F(3, 80) = 1.63$, $p = .19$. However, TAU in this trial may not reflect SEL provision in other countries, which we note as a study limitation in the main text.

Measures

Outcomes. Given our focus on depression prevention, our machine learning models predicted change in depressive symptoms from baseline to post-intervention, as measured by the Center for Epidemiologic Studies Depression Scale[6] (i.e., CESD_post − CESD_pre). Because our sample consisted of early adolescents (ages 11–13 at baseline)—a developmental period during which depressive symptoms often increase[7–12]—this trend was evident in our data. At the school level, where scores were averaged across students, 97.6% of schools (82 out of 84) showed higher mean CESD scores post-intervention. At the individual student level, 58.9% of students who completed both assessments (4451/7555) showed increased depressive symptoms

over time, with a modest mean increase of 3.44 CESD points (Cohen's $d = 0.31$) (see **Supplemental Figure 1** for school-level trends in depression change over time across all timepoints).

In the main trial[3], social–emotional-behavioural functioning (Strengths and Difficulties Questionnaire; SDQ Youth Self-Report Version[13]) and self-reported well-being (Warwick-Edinburgh Mental Well-being Scale; WEMBS[14]) were also co-primary outcomes. Thus, we also re-run analyses including these two variables as outcomes. Missingness was relatively low for post-intervention outcome variables (10% for CESD, SDQ, and WEMBS).

Original Analysis Plans and Deviations. In our pre-publication form submitted to the MYRIAD Publication Committee in September 2022, we prespecified analyses focused on the three co-primary outcomes from the original parent trial (CES-D, SDQ, WEMWBS) using a single train/test split design. The revised manuscript retains these core elements but incorporates several changes in response to reviewer feedback. Specifically, in response to an anonymous reviewer, we replaced the single train/test split with a nested 10-fold cross-validation procedure.

In terms of modeling frameworks, while our pre-publication plan considered a broader set of machine learning methods (elastic net regression, random forest, support vector machines, and an ensemble approach), in our manuscript we focused on two complementary approaches: elastic net regression (one of the models originally planned), a penalized regression that models additive, linear effects and has shown promise in prior treatment prediction work (including from our group[15,16]), and causal forests, a newer extension of random forest (the latter included in our pre-publication plan) explicitly designed to detect heterogeneous treatment effects by capturing non-linearities and higher-order interactions. Our adoption of causal forests was informed by Dr. Ronald Kessler's research team, who have applied this approach in recent and ongoing work[17] and in an ongoing collaboration with their group. To be clear, we became aware of causal forest

after we had submitted the pre-publication form, and we never ran a support vector machine model as we decided elastic net and causal forest were sufficient. Full hyperparameters and reproducible code are documented in the Supplement/OSF (https://osf.io/4ftq7/files). In the main text we note the limitation that the analysis plan was not formally preregistered on a public platform (e.g., OSF).

Analytic Approach

To estimate which individuals were likely to benefit more from SBMT versus TAU, we computed Personalized Advantage Index (PAI)[15,18–22] scores using two complementary machine learning approaches: (1) Elastic Net Regression (ENR) implemented via the glmnet package (version 4.1-8)[23] and (2) Causal Forests (CF) implemented via the grf package (version 2.3.2).[24] All analyses were implemented in R.

To compute PAI scores, we subtracted each individual's predicted CESD outcome score under SBMT from their predicted score under TAU. This yielded a PAI score for each participant, where positive values indicate an individual is expected to benefit more from SBMT, and negative values indicate they are expected to benefit more from TAU. The magnitude of the PAI score reflects the expected strength of this advantage—larger absolute values correspond to greater predicted differences in benefit between the two interventions. See Supplemental Table S1 for the list of predictor variables we selected for inclusion.

Elastic Net Regression (ENR). Elastic Net is a regularized variant of linear regression that combines L1 (lasso) and L2 (ridge) penalties to enhance predictive accuracy while accounting for multicollinearity. The inclusion of L1 penalty enables automatic variable selection when fitting the ENR. The ENR model included all main effects of baseline variables, a binary intervention group indicator (SBMT vs. TAU), and all group × predictor interaction terms to capture differences in the response profiles in the two treatment arms. Model training was

conducted using the *caret::train* function with repeated 10-fold cross-validation (10 repeats) and a hyperparameter tuning grid (tuneLength = 20) to identify optimal alpha and lambda values.

Causal Forest (CF). Causal Forest is a newer tree-based ensemble method that directly estimates heterogeneous treatment effects (HTEs), that is, how the effect of a treatment varies with the baseline covariates.[24] Unlike ENR, CF is capable of automatically detecting non-linear relationships and higher-order interactions without having to explicitly specify them in a model. CF works by building many decision trees, each designed to partition the data into subgroups where the difference in outcomes between treated and untreated individuals is most informative. These trees are then aggregated to produce stable and *individualized* treatment effect estimates. We first ran a preliminary CF with all baseline covariates included and selected the variables with a relative importance larger than 0.2 x mean importance (or the top 10 if fewer variables met this threshold).[17] These variables were then used in the final CF fit to estimate the HTEs.[17] We used the *causal_forest* function from the grf package, with 3,000 trees and the built-in hyperparameter tuning procedure, which selects the optimal configurations of the seven hyperparameters: (1) the fraction of data used to build each tree, (2) the number of variables randomly sampled at each split, (3) the minimum number of observations per leaf node, (4) the fraction of data reserved for honest splitting, (5) whether trees should be pruned, (6) the allowable level of imbalance in splits, and (7) the penalty strength on imbalanced splits. The tuning process utilized smaller forests with 500 trees and was repeated for 10 times for stability. We considered a total of 1000 candidate configurations of the hyperparameters. Expected outcomes given the covariates, marginalized over treatment assignment, were estimated using the *SuperLearner* with the following three base learners: 1) ridge regression, 2) lasso regression, 3) elastic net regression with a 0.5 weights for both the L1 and L2 penalties, 4) xgboost, and 5) random forest from the SuperLearner package.[25] Each was preceded by two-stage screeners: (i) a

correlation filter that removes highly collinear predictors ($|r| \geq 0.8$), followed by (ii) either CV-lasso capped at 10/20/30 non-zero coefficients or random-forest retaining the top 10/20/30 variables by corrected impurity importance. This yields multiple screened variants per base learner to support robust ensembling.

Prediction performance and goodness-of-fit of the PAI models. We assessed the general predictive performance of both machine learning approaches to evaluate their capability to predict depression outcomes from baseline characteristics. However, the specific architectures of these models necessitate different approaches. For Causal Forest, the model outputs individual-level conditional treatment effect estimates (CATEs) rather than marginal outcome predictions, making standard predictive metrics less informative. Instead, we report the results from the Best Linear Predictor (BLP) test.[26] This approach regresses the forest's predicted intervention effects against the observed outcomes in an auxiliary regression, with the resulting coefficients providing a calibration check. A well-calibrated forest should yield a slope close to one. In our analysis, we performed this check for each iteration of the nested cross-validation and reported the mean estimated slope. For the Elastic Net model, we evaluated performance by assessing the accuracy of predicted outcomes under the assigned intervention in the hold-out set. Performance metrics included R^2, root mean squared error (MSE), and mean absolute error (MAE). Critically, the ultimate test of predictive performance for intervention selection models is whether the algorithms successfully identify which individuals benefit from which intervention condition—this is directly assessed via the PAI analyses described below.

Generating Predicted Outcomes to Compute PAI Scores. For each individual, we generated (via nested cross-validation) two predictions: their expected factual and counterfactual outcomes if assigned to SBMT vs. TAU. For the ENR model, this was achieved by varying the binary intervention group indicator (SBMT vs. TAU) while holding all other covariates constant.

The PAI score was calculated as the difference between these predicted outcomes (TAU - SBMT), with positive scores indicating a greater expected benefit from SBMT. In contrast, the CF model directly outputs an estimate of this expected group difference (TAU – SBMT) for each individual, which we used as their PAI score.

Evaluation of Treatment Recommendations. We tested whether PAI scores moderated intervention group differences in depression outcome (i.e., Intervention Group x PAI interaction), providing a formal statistical test of whether individuals who were assigned to their algorithm-indicated intervention (as identified by the model) had better outcomes.

While the above analyses treat PAI scores as a *continuous* moderator, we next examined whether binarizing the sample into model-identified subgroups—those predicted to benefit more from SBMT (PAI > 0) or TAU (PAI < 0) —would similarly yield meaningful differences in observed outcomes. This approach allows for a more direct test of the models' clinical utility: If a student is predicted to do better with SBMT (or TAU), do they in fact experience better outcomes if they were randomly assigned to that intervention? We conducted this subgroup analysis separately for each model to assess how well their binary intervention recommendations mapped onto actual depressive symptom change.

The above depression (CESD) analyses were repeated using two additional outcome measures: SDQ and WEMWBS.

Decision curve analysis. Since the individual treatment effect of mindfulness over TAU is not observed, the conventional decision curve analysis is not directly applicable in our context. We instead performed Qini curve analysis, which is tailored towards cost-benefit analysis of treatment allocation polices.[27] Qini curve analysis provides a way to evaluate individualized treatment assignment rules by quantifying their ability to prioritize individuals who derive the greatest benefit from treatment. Suppose we have estimated individual treatment effects $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ for

$\boldsymbol{i = 1, \ldots, n}$. We first sort individuals in descending order of $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$. For a given fraction $\boldsymbol{q} \in [\boldsymbol{0}, \boldsymbol{1}]$, define the top-$\boldsymbol{qn}$ subgroup as those with the $\boldsymbol{qn}$ largest predicted effects. $\boldsymbol{q}$ here can also be interpreted as the average cost of a treatment assignment rule if we assume that the cost of mindfulness intervention is a constant across all subjects. To estimate the realized incremental gain, we use a doubly robust score for each unit,

$$\boldsymbol{\phi_i = \hat{\mu}_1(X_i) - \hat{\mu}_0(X_i) \; + \; \frac{W_i}{\hat{e}(X_i)}(Y_i - \hat{\mu}_1(X_i)) \; - \; \frac{1 - W_i}{1 - \hat{e}(X_i)}(Y_i - \hat{\mu}_0(X_i))},$$

where $\hat{\boldsymbol{\mu}}_{\boldsymbol{w}}(\boldsymbol{X_i})$ are estimated conditional mean outcome for treatment $\boldsymbol{w} \in \{\boldsymbol{0}, \boldsymbol{1}\}$ and $\hat{\boldsymbol{e}}(\boldsymbol{X_i})$ is the estimated propensity score. We used the package *grf* to compute this score directly. The cumulative incremental gain function is then

$$\boldsymbol{G(q) = \frac{1}{qn} \sum_{i=1}^{qn} \phi_i},$$

which captures the average treatment effect among those individuals most prioritized by the model. The Qini curve is obtained by plotting $\boldsymbol{G(q)}$ against the fraction $\boldsymbol{q}$ of the population treated. A perfectly informative model would yield a curve that rises steeply, indicating that those with the highest predicted $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ indeed realize the greatest benefit, while a non-informative model produces a near-linear curve. Note that for a personalized treatment assignment rule, there is typically an upper bound on $\boldsymbol{q}$, since individuals with negative estimated treatment effects would not be assigned to mindfulness. This upper bound corresponds to the proportion of observations with positive $\hat{\boldsymbol{\tau}}_{\boldsymbol{i}}$ in the dataset. Our Qini curve is derived from the nested cross-validation procedure. For each fold, we used GRF to estimate the doubly robust scores $\boldsymbol{\phi_i}$ for subjects in that fold, while the corresponding individual treatment effect estimates were obtained

from the PAI score model trained on the data excluding that fold. A single Qini curve analysis was performed on the merged fold-specific results.

Results

Outcomes Among Model-Identified Subgroups. For the Causal Forest model, individuals predicted to benefit more from SBMT (i.e., those with PAI scores > 0) demonstrated significantly better depression outcomes when randomly assigned to SBMT compared to TAU, but the magnitude of the effect was negligible (d = 0.07, p = .007; **Figure 4, left panel**). In contrast, among those predicted to benefit more from TAU (PAI < 0), no significant difference in outcomes was observed between the two groups (d = -0.01, p = .815).

A similar pattern emerged for the Elastic Net model. Individuals predicted to benefit more from TAU (PAI < 0) showed significantly better outcomes when assigned to TAU compared to SBMT, but the magnitude of the effect was negligible (d = 0.08, p = .004; **Figure 2, right panel**). For those predicted to benefit more from SBMT (PAI > 0), outcomes did not significantly differ by treatment assignment (d = 0.00, p = .997).

Enhanced Effects in High-Confidence (larger PAI) Subgroup? Given that larger absolute PAI scores indicate a greater expected magnitude of benefit from a specific intervention, we conducted an additional subgroup analysis focusing on participants with stronger model recommendations. We identified a subgroup for whom both the Causal Forest and Elastic Net models yielded more extreme predictions—specifically, participants with absolute PAI scores above the median ($|PAI| > 50^{th}$ percentile) for each model. Within this subgroup, effects of intervention matching were only slightly more pronounced. Youth assigned to SBMT who were strongly indicated for SBMT demonstrated a negligible benefit (CF: d = 0.12; ENR: d = 0.13).

Additional Outcomes. We re-ran both the Elastic Net and Causal Forest models using the SDQ and WEMWBS outcomes (i.e., recomputing PAI scores reflecting predicted post-intervention outcomes for the latter two measures). Of the four Group × PAI interaction tests conducted (2 models × 2 outcomes), none were statistically significant.

Reduced Set of Predictors: In response to an anonymous review, we reran the primary analysis predicting change in CESD with a reduced set of 9 predictor variables (derived from a prior paper[5]): including six student-level variables (pre-intervention scores on the 3 outcome measures, age, sex, and ethnicity) and three school-level characteristics: urbanicity (rural vs. urban), level of school deprivation (% of students eligible for free school meals) and the SEL ethos (composite measure of school quality, teacher-rated school climate and the school's SEL provision). The Group × PAI interaction was statistically significant for the Elastic Net model ($p = .028$) but not Causal Forest. Among youth predicted by the elastic net model to benefit more from TAU, those who were randomly assigned to TAU showed significantly better (but clinically negligible) depression outcomes relative to those assigned to SBMT ($d = 0.10$, $p = 0.034$). There were no differences for other comparisons.

Decision curve. The Qini curve analysis compared the treatment assignment rules generated by the Causal Forest and Elastic Net models against a random baseline, where a proportion of patients were randomly selected to receive the mindfulness intervention. As shown in the **Supplementary Figure 2,** both the Causal Forest model and the Elastic Net model achieved greater cumulative benefit across a range of treatment fractions (i.e., average costs) compared to the random assignment rule. When the treatment fraction is below 0.5, the Elastic Net slightly outperforms the Causal Forest, though there are regions where the Causal Forest provides higher benefits. Above the treatment fraction of 0.5, however, the Causal Forest shows substantial improvement over the Elastic Net, with its Qini curve reaching a higher plateau.

Overall, these findings suggest that there is no clear advantage when resources are limited and only individuals with the highest predicted treatment effects can be assigned to mindfulness. In contrast, when it is feasible to assign a larger proportion of individuals, the Causal Forest produces markedly greater overall benefit.

References:

1. Kuyken W, Nuthall E, Byford S, et al. The effectiveness and cost-effectiveness of a mindfulness training programme in schools compared with normal school provision (MYRIAD): study protocol for a randomised controlled trial. *Trials*. 2017;18(1):194. doi:10.1186/s13063-017-1917-4

2. Ford T, Degli Esposti M, Crane C, et al. The Role of Schools in Early Adolescents' Mental Health: Findings From the MYRIAD Study. *J Am Acad Child Adolesc Psychiatry*. 2021;60(12):1467-1478. doi:10.1016/j.jaac.2021.02.016

3. Kuyken W, Ball S, Crane C, et al. Effectiveness and cost-effectiveness of universal school-based mindfulness training compared with normal school provision in reducing risk of mental health problems and promoting well-being in adolescence: the MYRIAD cluster randomised controlled trial. *Evid Based Ment Health*. 2022;25(3). doi:10.1136/ebmental-2021-300396

4. Montero-Marin J, Nuthall E, Byford S, et al. Update to the effectiveness and cost-effectiveness of a mindfulness training programme in schools compared with normal school provision (MYRIAD): study protocol for a randomised controlled trial. *Trials*. 2021;22(1):254. doi:10.1186/s13063-021-05213-9

5. Montero-Marin J, Allwood M, Ball S, et al. School-based mindfulness training in early adolescence: what works, for whom and how in the MYRIAD trial? *BMJ Ment Health*. 2022;25(3):117-124. doi:10.1136/ebmental-2022-300439

6. Radloff LS. The CES-D Scale A Self-Report Depression Scale for Research in the General Population. *Appl Psychol Meas*. 1977;1(3):385-401. doi:10.1177/014662167700100306

7. McGrath JJ, Al-Hamzawi A, Alonso J, et al. Age of onset and cumulative risk of mental disorders: a cross-national analysis of population surveys from 29 countries. *Lancet Psychiatry*. 2023;10(9):668-681. doi:10.1016/S2215-0366(23)00193-1

8. Avenevoli S, Swendsen J, He JP, Burstein M, Merikangas KR. Major depression in the national comorbidity survey–adolescent supplement: Prevalence, correlates, and treatment. *J Am Acad Child Adolesc Psychiatry*. 2015;54(1):37-44.e2. doi:10.1016/j.jaac.2014.10.010

9. Weissman MM, Warner V, Wickramaratne P, Moreau D, Olfson M. Offspring of Depressed Parents: 10 Years Later. *Arch Gen Psychiatry*. 1997;54(10):932-940. doi:10.1001/archpsyc.1997.01830220054009

10. Alloy LB, Abramson LY. The adolescent surge in depression and emergence of gender differences: A biocognitive vulnerability-stress model in developmental context. In: *Adolescent Psychopathology and the Developing Brain: Integrating Brain and Prevention Science*. Oxford University Press; 2007:284-312. doi:10.1093/acprof:oso/9780195306255.003.0013

11. Hankin BL, Abramson LY, Moffitt TE, Silva PA, McGee R, Angell KE. Development of depression from preadolescence to young adulthood: Emerging gender differences in a 10-year longitudinal study. *J Abnorm Psychol*. 1998;107(1):128-140. doi:10.1037/0021-843X.107.1.128

12. Costello EJ, Copeland W, Angold A. Trends in psychopathology across the adolescent years: What changes when children become adolescents, and when adolescents become adults? *J Child Psychol Psychiatry*. 2011;52(10):1015-1025. doi:10.1111/j.1469-7610.2011.02446.x

13. Goodman R. Psychometric properties of the strengths and difficulties questionnaire. *J Am Acad Child Adolesc Psychiatry*. 2001;40(11):1337-1345. doi:10.1097/00004583-200111000-00015

14. Tennant R, Hiller L, Fishwick R, et al. The Warwick-Edinburgh Mental Well-being Scale (WEMWBS): development and UK validation. *Health Qual Life Outcomes*. 2007;5(1):63. doi:10.1186/1477-7525-5-63

15. Webb CA, Hirshberg MJ, Davidson RJ, Goldberg SB. Personalized Prediction of Response to Smartphone-Delivered Meditation Training: Randomized Controlled Trial. *J Med Internet Res*. 2022;24(11):e41566. doi:10.2196/41566

16. Webb CA, Cohen ZD, Beard C, Forgeard M, Peckham AD, Björgvinsson T. Personalized prognostic prediction of treatment outcome for depressed patients in a naturalistic psychiatric hospital setting: A comparison of machine learning approaches. *J Consult Clin Psychol*. 2020;88(1):25-38. doi:10.1037/ccp0000451

17. Zainal NH, Bossarte RM, Gildea SM, et al. Developing an individualized treatment rule for Veterans with major depressive disorder using electronic health records. *Mol Psychiatry*. 2024;29(8):2335-2345. doi:10.1038/s41380-024-02500-0

18. Cohen ZD, DeRubeis RJ, Hayes R, et al. The development and internal evaluation of a predictive model to identify for whom Mindfulness-Based Cognitive Therapy (MBCT) offers superior relapse prevention for recurrent depression versus maintenance antidepressant medication. *Clin Psychol Sci J Assoc Psychol Sci*. 2023;11(1):59-76. doi:10.1177/21677026221076832

19. Cohen ZD, DeRubeis RJ. Treatment selection in depression. *Annu Rev Clin Psychol*. 2018;14(1):209-236. doi:10.1146/annurev-clinpsy-050817-084746

20. Webb CA, Swords CM, Lawrence HR, Hilt LM. Which adolescents are well-suited to app-based mindfulness training? A randomized clinical trial and data-driven approach for personalized recommendations. *J Consult Clin Psychol*. 2022;90:655-669. doi:10.1037/ccp0000763

21. Webb CA, Trivedi MH, Cohen ZD, et al. Personalized prediction of antidepressant v. placebo response: evidence from the EMBARC study. *Psychol Med*. Published online July 2018:1-10. doi:10.1017/S0033291718001708

22. DeRubeis RJ, Cohen ZD, Forand NR, Fournier JC, Gelfand LA, Lorenzo-Luaces L. The Personalized Advantage Index: Translating Research on Prediction into Individualized Treatment Recommendations. A Demonstration. *PLOS ONE*. 2014;9(1):e83875. doi:10.1371/journal.pone.0083875

23. Friedman JH, Hastie T, Tibshirani R. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 2010;33(1):1-22.

24. Athey S, Tibshirani J, Wager S. Generalized random forests. *Ann Stat*. 2019;47(2):1148-1178. doi:10.1214/18-AOS1709

25. Polley EC, LeDell E, van der Laan MJ. SuperLearner: Super Learner Prediction (Version 2.0-23) [R Package]. Published online 2016.

26. Athey S, Imbens G, Pham T, Wager S. Estimating Average Treatment Effects: Supplementary Analyses and Remaining Challenges. *Am Econ Rev*. 2017;107(5):278-281. doi:10.1257/aer.p20171042

27. Sverdrup E, Wu H, Athey S, Wager S. Qini Curves for Multi-Armed Treatment Rules. *J Comput Graph Stat*. 2025;34(3):948-960. doi:10.1080/10618600.2024.2418820

28. Hinze V, Ford T, Crane C, et al. Does depression moderate the relationship between pain and suicidality in adolescence? A moderated network analysis. *J Affect Disord*. 2021;292:667. doi:10.1016/j.jad.2021.05.100

29. Department for Education. https://assets.publishing.service.gov.uk/media/67bde335729cf14d438d76c2/Secondary_accountability_measures_-_guide_for_maintained_secondary_schools__academies_and_free_schools.pdf. 2025. Accessed August 8, 2025. https://assets.publishing.service.gov.uk/media/67bde335729cf14d438d76c2/Secondary_accountability_measures_-_guide_for_maintained_secondary_schools__academies_and_free_schools.pdf

Table S1

Predictors included in the machine learning models

Predictor	**Definition**	**Variable level**
CESD	Center for Epidemiologic Studies Depression Scale (CESD) total score at baseline	Student-level
WEM	Warwick-Edinburgh Mental Well-being Scale (WEMWBS) total score at baseline	Student-level
SDQ	Strengths and Difficulties Questionnaire (SDQ) total difficulties score at baseline	Student-level
Age	Age based on the date of birth at baseline, rounded to nearest year	Student-level
MinorityinSchool	Pupils ethnicity is a minority in school at baseline	Student-level
Sex_female_c	Female yes or no (derived from gender question at baseline with response options: female, male, other, and prefer not to say)	Student-level

Predictor	Definition	Variable level
Sex_other_c	Sex “other” yes or no	Student-level
Sex_notsay_c	Sex “prefer not to say” yes or no	Student-level
SuicidalityTotal	Suicidality all variables combined. See 3 suicidality items in Hinze et al. (2021). Response options: no, yes, prefer not to say, and unclear.	Student-level
Suicidality_TotalNo	Total of “no” responses across 3 suicidality items	Student-level
Suicidality_TotalYes	Total of “yes” responses across 3 suicidality items	Student-level
Suicidality_TotalPnts	Total of “prefer not to say” responses across 3 suicidality items	Student-level
CAMM	Child and Adolescent Mindfulness Measure (CAMM) total score	Student-level
PDS_PubertyCat_male_scaled	Male puberty categories: 0 (prepuberty) – 4 (postpuberty)	Student-level

Predictor	Definition	Variable level
PDS_PubertyCat_female_scaled	Female puberty categories: 0 (prepuberty) – 4 (postpuberty)	Student-level
Self_01	Life not worth living: ”Have you thought that life was not worth living, or that you would be better off dead?”. See Hinze et al. [28]	Student-level
Self_02	Thought about harming yourself: “Have you thought seriously about trying to harm yourself in some way (for example by cutting yourself or taking an overdose of pills or other medication)?”. See Hinze et al. [28]	Student-level
Self_03	Deliberately harmed yourself: “Have you actually, deliberately harmed yourself in some way (for example by cutting yourself or taking an overdose of pills or other medication)?”. See Hinze et al. [28]	Student-level
Risk_03	E-cigarette smoking habit (single item) Which of these best describes you?	Student-level

Predictor	Definition	Variable level
	Think about times when you may have a puff or two as well as using whole electronic cigarettes Tick one box only - I have never tried electronic cigarettes - I have used electronic cigarettes only once or twice - I used to use electronic cigarettes, but I don't now - I sometimes use electronic cigarettes, but I don't use them every week - I use electronic cigarettes regularly, once a week or more	
Risk_04	Alcohol habit (single item) Which of the following apply to you. I drink alcohol: Tick one box only - never - only a few times a year - about once a month - about once a fortnight - about once a week - about twice a week - every day or almost every day	Student-level

Predictor	Definition	Variable level
SDQ_EMO	SDQ emotional problems subscale score at baseline	Student-level
SDQ_COND	SDQ conduct problems subscale score at baseline	Student-level
SDQ_HYP	SDQ hyperactivity subscale score at baseline	Student-level
SDQ_PEER	SDQ peer problems subscale score at baseline	Student-level
SDQ_PRSOC	SDQ prosocial subscale score at baseline	Student-level
CHU9D_utility	Total score of the Child Health Utility (CHU9D)	Student-level
Pupil_SCCS_UpdatedTotal	The Alaska School Climate and Connectedness Survey (SCCS) - pupil total score across the subscales School Leadership and Involvement, Respectful Climate, Peer Climate, and Caring Adults) updated	Student-level
RCADS_GAD_T	Revised Children's Anxiety and Depression	Student-level

Predictor	Definition	Variable level
	Scale (RCADS), subscale generalised anxiety disorder T-score (see https://rcads.ucla.edu)	
RCADS_OCD_T	RCADS, subscale obsessive-compulsive disorder T-score	Student-level
RCADS_PD_T	RCADS, subscale panic disorder T-score	Student-level
RCADS_SAD_T	RCADS, subscale separation anxiety disorder T-score	Student-level
RCADS_SOC_T	RCADS, subscale social phobia T-score	Student-level
RCADS_Total	RCADS total T-score	Student-level
Brief2_Inhibit	Behavior Rating Inventory of Executive Function, Second Edition (BRIEF2) inhibit subscale score	Student-level
Brief2_SelfMonitor	BRIEF2 self-monitor subscale score	Student-level
Brief2_Shift	BRIEF2 shift subscale score	Student-level

Predictor	**Definition**	**Variable level**
Brief2_EmotionalControl	BRIEF2 emotional control subscale score	Student-level
Brief2_TaskCompletion	BRIEF2 task completion subscale score	Student-level
Brief2_WorkMemory	BRIEF2 working memory subscale score	Student-level
Brief2_Plan	BRIEF2 planning subscale score	Student-level
class	Most likely profile membership based on a latent profile analysis of baseline characteristics (for details, see [5])	Student-level
AP_BF2_01	How well the teacher knows the student (1, Not well; 2, Moderately well; 3, Very well)	Student-level (but teacher-rated)
AP_BF2_02	How long the teacher knows the student	Student-level (but teacher-rated)
AP_Breif2_total	BRIEF2 total score	Student-level (but teacher-rated)
AP_SDQ_EMO	SDQ emotional problems subscale score	Student-level (but teacher-rated)

Predictor	Definition	Variable level
AP_SDQ_COND	SDQ conduct problems subscale score	Student-level (but teacher-rated)
AP_SDQ_HYP	SDQ hyperactivity subscale score	Student-level (but teacher-rated)
AP_SDQ_PEER	SDQ peer problems subscale score	Student-level (but teacher-rated)
AP_SDQ_PRSOC	SDQ prosocial subscale score	Student-level (but teacher-rated)
AP_SDQ_TOT	SDQ total score	Student-level (but teacher-rated)
AP_SDQ_IMPACT	SDQ impact score	Student-level (but teacher-rated)
C_Teacher_Overall	MBI:TAC Overall teaching competency rating for teaching mindfulness lessons (externally rated) (1, Incompetent; 2, Beginner; 3, Advanced Beginner; 4, Competent; 5, Proficient; 6, Advanced)	Teacher-level
CohortID_c	Cohort recoded (binary variable representing whether cohort was recruited in the academic years 2016/2017 or 2017/2018)	School-level

Predictor	Definition	Variable level
CrossBinaryUrbanRural	Urban/rural classification at baseline. See Ford et al. [2]	School-level
PercentageEligibleFreeSchoolMeal	Percentage of pupils eligible for free school meals at baseline. See Ford et al. [2]	School-level
Baseline_SEL_Factor_scores	Social-emotional learning factor scores at baseline. See Ford et al. [2]	School-level
NumOfPupilsWS	Number of pupils in whole school at baseline. See Ford et al. [2]	School-level
PupilTeacherRatio	Pupil:Teacher ratio at baseline. See Ford et al. [2]	School-level
IncomeDecile	Income decile at baseline	School-level
EmploymentDecile	Employment decile at baseline	School-level
HTAllocationPreference	Consenting headteacher's intervention allocation preference at baseline	School-level

Predictor	Definition	Variable level
	(.00, No Preference; 1.00, Mindfulness; 2.00, Teaching as Usual)	
TypeOfSchool	Type of school at baseline (1, State-funded secondary - Comprehensive (secondary); 2, State-funded secondary- Modern (secondary); 3, State-funded secondary - Selective (grammar); 4, State-funded secondary - Unknown; 5, Independent school).	School-level
SchoolSex_c	School sex recoded (mixed or single-sex [female only]) at baseline. See Ford et al. [2]	School-level
InspectionMappingRandomisation	Inspection mapping (school quality rating) at randomization (1.00, Outstanding; 2.00, Good; 3.00 Satisfactory; 4.00, Requires Improvement; 5.00, Inadequate)	School-level
SchoolProgress	Progress 8 at baseline. See page 12 in [29]	School-level
PercentageWhiteBritishEthnicOrig	Percentage of students with white British ethnic	School-level

Predictor	Definition	Variable level
	origin within the school at baseline	
PercentageLocalEthnicity_White	Percentage of local ethnicity "White" in the schools local area at baseline	School-level
Pupil_SCCS_SLSI_SchLev	SCCS School Leadership and Involvement subscale (school-level aggregate of student ratings)	School-level
Pupil_SCCS_RC_SchLev	SCCS Respectful Climate subscale (school-level aggregate of student ratings)	School-level
Pupil_SCCS_PCS_SchLev	SCCS Peer Climate subscale (school-level aggregate of student ratings)	School-level
Pupil_SCCS_CAS_SchLev	SCCS Caring Adults subscale (school-level aggregate of student ratings)	School-level
Pupil_SCCS_Total_SchLev	Total Score of the SCCS (school-level aggregate of student ratings)	School-level
SCCS_Teacher_SLI_SchLev	SCCS School Leadership and Involvement subscale	School-level

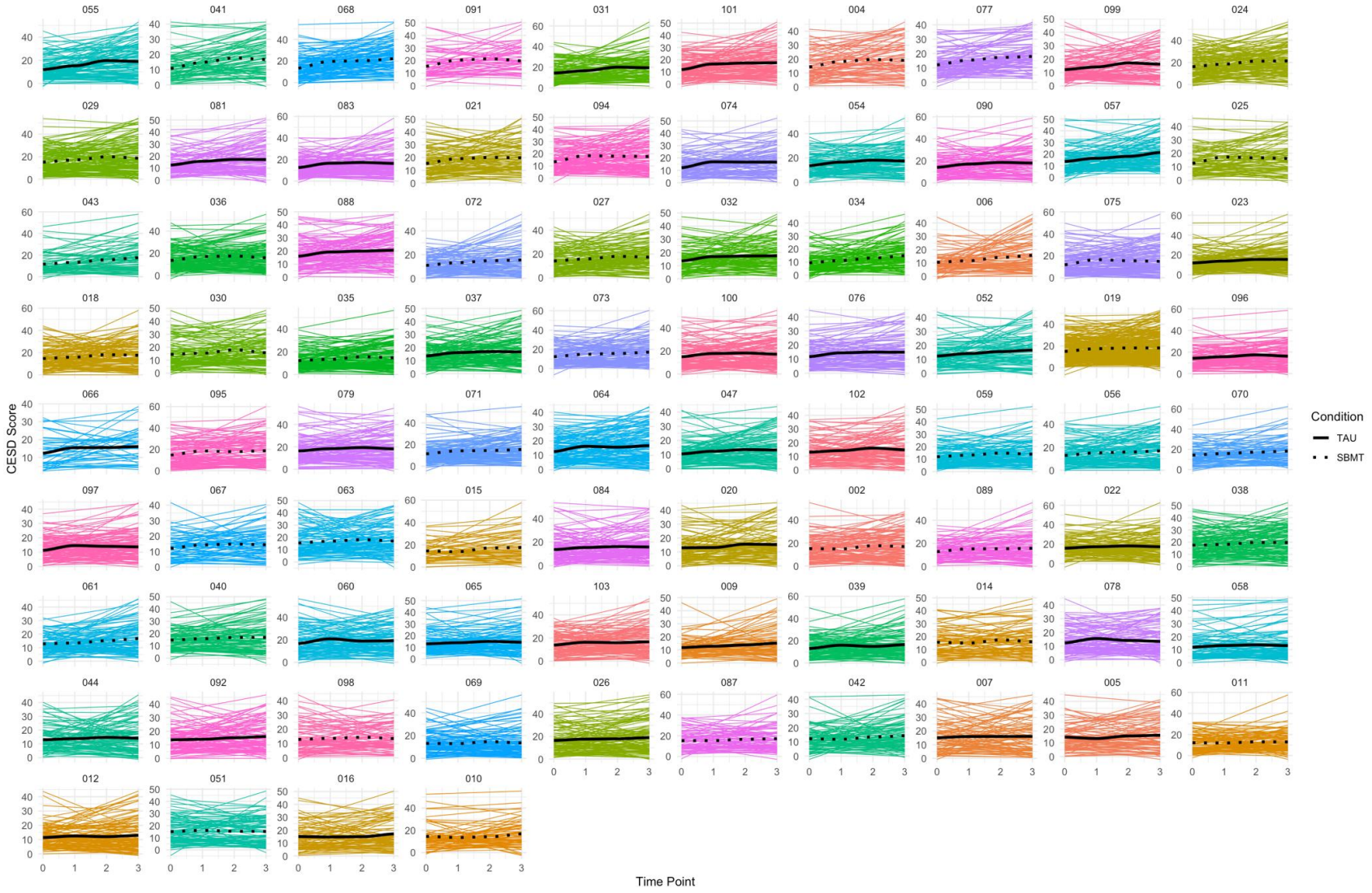
Predictor	Definition	Variable level
	(school-level aggregate of teacher ratings)	
SCCS_Teacher_SA_SchLev	SCCS Staff Attitudes subscale (school-level aggregate of teacher ratings)	School-level
SCCS_Teacher_RC_SchLev	SCCS Respectful Climate subscale (school-level aggregate of teacher ratings)	School-level
SCCS_Teacher_Total_SchLev	SCCS total score across the subscale School Leadership and Involvement, Staff Attitudes, and Respectful Climate (school-level aggregate of teacher ratings)	School-level

Table S2. Baseline characteristics of students by PAI Group

	Causal Forest		Elastic Net	
	SBMT	**TAU**	**SBMT**	**TAU**
Gender, n (%)				
Female	3391 (52.7%)	1118 (62.5%)	2735 (54.2%)	1774 (55.9%)
Male	2911 (45.3%)	636 (35.5%)	2208 (43.7%)	1339 (42.2%)
Age, mean (SD)	11.7 (0.6)	11.7 (0.6)	11.7 (0.6)	11.7 (0.6)
Ethnicity—White, n (%)	4971 (77.5%)	1231 (69.2%)	3838 (76.2%)	2364 (74.9%)
Depression (CES-D), mean (SD)	15.4 (10.9)	16.3 (11.5)	15.6 (11.1)	15.5 (10.9)
Social-emotional-behavioral functioning (SDQ), mean (SD)	12.3 (6.6)	12.7 (6.8)	12.4 (6.6)	12.4 (6.6)
Well-being (WEMWBS), mean (SD)	49.2 (9.0)	48.6 (9.2)	49.1 (9.0)	49.1 (9.1)

Notes. Also see Table 1 in main text for differences by intervention group. Pre-intervention CES-D, Center for Epidemiological Studies for Depression Scale; PSHE, personal, social, health and economic; SDQ, Strengths and Difficulties Questionnaire; SEL ethos, school social–emotional learning ethos; WEMWBS, Warwick- Edinburgh Mental Well- Being Scale. For each model (Causal Forest and Elastic Net), we generated descriptive statistics using the *tableone* package in R. For categorical variables, percentages represent the proportion of students within each PAI subgroup (within columns) who belong to that category.

Supplemental Fig 1. Trends in depression scores in every school in the MYRIAD trial



Supplemental Fig 2. QINI curves of the treatment assignment rules constructed from the PAIs of the Causal Forest and Elastic Net. The dashed line indicates a baseline rule that randomly assigns subjects to the mindfulness arm.

